Dual-Constraint Contrastive Completion: Bridging Local Precision and Global Integrity in 3D Point Cloud Reconstruction

Donald Martin * and Blake Bowman

Posted Date: 3 April 2026

doi: 10.20944/preprints202604.0233.v1

Keywords: point cloud completion; contrastive learning; asymmetric chamfer distance; self-supervised learning; 3D point clouds

 doi:10.20944/preprints202604.0233.v1

Article

Dual-Constraint Contrastive Completion: Bridging Local Precision and Global Integrity in 3D Point Cloud Reconstruction

Donald Martin * and Blake Bowman

Jefferson University, USA
* Correspondence: rbrown53@my.ncu.edu.jm

Abstract

Incomplete 3D point clouds present a significant challenge in diverse applications due to sensor limitations and occlusions. Existing methods often struggle to balance local detail accuracy and global structural integrity, frequently yielding artifacts or distortions due to reliance on symmetric Chamfer Distance or unguided contrastive losses. We propose Dual-Constraint Contrastive Completion (DCCC), an end-to-end framework integrating asymmetric weighted Chamfer distance with multi-granularity contrastive learning. DCCC utilizes an encoder-decoder backbone with Mamba layers for efficient feature extraction. Central is the Asymmetric Contrastive Chamfer Loss (ACCL), decoupling local precision and global integrity objectives into distinct contrastive components, optimized via dynamic asymmetric weighting. A Self-Supervised Structural Guidance (SSG) module further learns coarse structural priors directly from incomplete inputs, reducing annotation reliance and improving robustness. Extensive experiments on benchmark datasets demonstrate DCCC's superior performance. DCCC achieves best-in-class results across critical metrics, significantly enhancing structural completeness and fine-grained accuracy in diverse settings, including real-world scenarios and high sparsity.

Keywords: point cloud completion; contrastive learning; asymmetric chamfer distance; self-supervised learning; 3D point clouds

1. Introduction

Three-dimensional (3D) point clouds, serving as discrete representations of objects and scenes in the digital world, play a pivotal role in diverse applications such as autonomous driving, robotics, and virtual reality [1]. However, point cloud data acquired from real-world sensors often suffers from incompleteness or sparsity due to various factors like occlusion, sensor limitations, or environmental interference [2], a common challenge also observed in 3D medical imaging and reconstruction techniques [3]. This inherent challenge severely constrains the performance of subsequent 3D tasks. Consequently, **3D point cloud completion**, which aims to accurately reconstruct missing regions from incomplete point clouds, stands as a fundamental and critical challenge in the field of 3D vision [4].

Existing research has made significant strides in point cloud completion, primarily focusing on two core directions: the design and improvement of loss functions and the introduction of advanced learning paradigms [5,6]. Chamfer Distance (CD) is widely adopted as the most common distance metric for point cloud completion, appreciated for its computational efficiency and geometric intuitiveness [7]. Nevertheless, the symmetric weighting mechanism of standard CD often struggles to balance the **local detail accuracy** and **global structural integrity** of generated point clouds, frequently leading to issues such as point accumulation, holes, or geometric distortion [8]. Recent works, such as FCD [9], have attempted to address this by introducing asymmetric weighting [6]. On another front, **contrastive learning**, a powerful self-supervised learning paradigm, has demonstrated its potential

 doi:10.20944/preprints202604.0233.v1

to enhance model representation capabilities and robustness across various visual tasks [10]. In the context of point cloud completion, contrastive learning has been applied in loss function design (e.g., InfoCD [11]) or at different hierarchical levels (e.g., Point Patches CL, Cluster+Instance CL [12]) to bolster local consistency and structural awareness. Furthermore, the rise of self-supervised learning has offered new avenues for reducing reliance on expensive and fully annotated complete point cloud data [13].

Despite the distinct focuses of these methods, a persistent challenge remains: how to organically combine the fine-grained control offered by Chamfer Distance (especially through asymmetric weighting) with the powerful representation capabilities of contrastive learning. The goal is to simultaneously optimize both local accuracy and global integrity while effectively leveraging the intrinsic structural information within point clouds. This research endeavors to bridge this gap by proposing a novel point cloud completion method that achieves high-fidelity and structurally consistent point cloud completion through a dual-constraint contrastive learning framework.

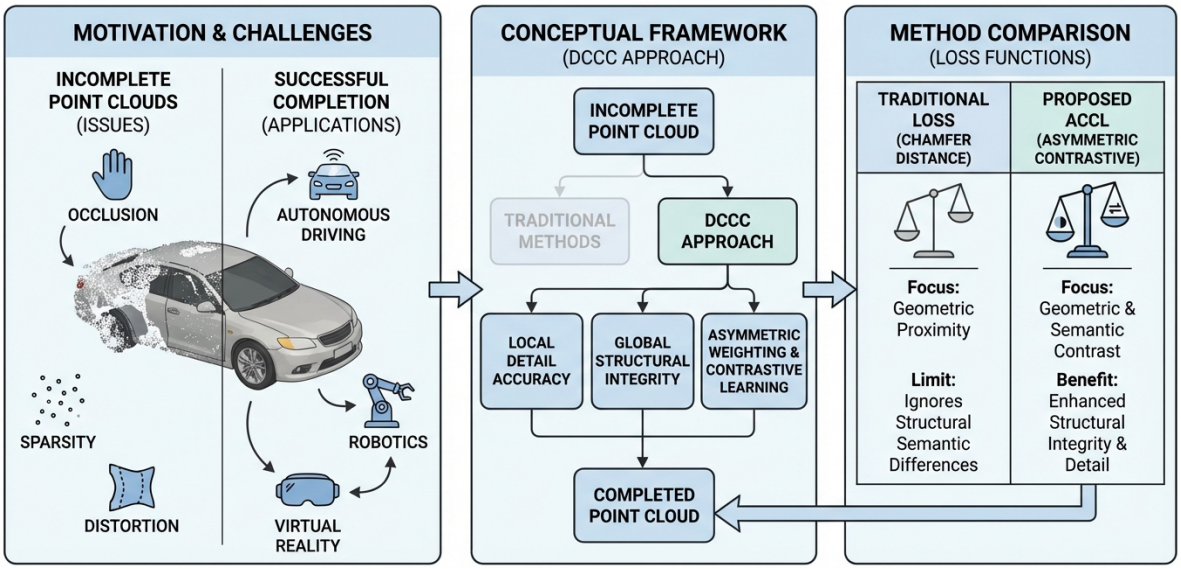

Figure 1. An illustrative overview of our Dual-Constraint Contrastive Completion (DCCC) framework. The figure highlights the prevalent challenges in 3D point cloud completion and its diverse applications, outlines the conceptual methodology of the DCCC approach emphasizing local detail, global structural integrity, and asymmetric contrastive learning, and provides a comparative perspective between traditional Chamfer Distance and our novel Asymmetric Contrastive Chamfer Loss (ACCL).

We propose **Dual-Constraint Contrastive Completion (DCCC)**, an end-to-end point cloud completion network that seamlessly integrates asymmetric weighted Chamfer distance with multi-granularity contrastive learning. DCCC's core idea is to explicitly separate and weight the learning objectives for local details and global structure, and then to utilize contrastive learning at different granularities to reinforce the achievement of these objectives. This approach ensures high fidelity and geometric consistency throughout the completion process. Our method features a robust encoder-decoder backbone that fuses the efficiency of Mamba architecture with the global modeling capabilities of Transformers (inspired by Simba [14]), effectively handling both long-range dependencies and local feature extraction in point clouds, drawing parallels with efficient information fusion strategies developed in other multi-modal learning domains [15]. At its heart lies the **Asymmetric Contrastive Chamfer Loss (ACCL)**, which decomposes the Chamfer Distance into local precision and global integrity components, each enhanced by a dedicated contrastive learning module. These modules are optimized with dynamic asymmetric weights to balance the learning progress. Additionally, a **Self-Supervised Structural Guidance** module is introduced to learn a coarse structural prior from incomplete inputs, further improving completion quality and reducing annotation reliance.

 doi:10.20944/preprints202604.0233.v1

To thoroughly evaluate DCCC's performance, we conduct extensive experiments on multiple standard datasets and utilize a comprehensive set of evaluation metrics. We leverage large-scale synthetic datasets such as **ShapeNet-55/ShapeNetPart** [16] and the **PCN Dataset** [17] to assess the model's generalization capabilities and performance under varying degrees of incompleteness. Furthermore, the **KITTI** [18] dataset, representing real-world autonomous driving scenarios, is used to validate the model's robustness and cross-domain generalization in completing LiDAR scans affected by occlusion. Our evaluation metrics include Chamfer Distance (ℓ_1 and ℓ_2) [7], Earth Mover's Distance (EMD) [7], Directed Chamfer Distance (DCD) [19], F-Score [20], and MMD [21], which collectively offer a comprehensive assessment of local accuracy, global shape integrity, and fine-grained geometric matching.

Our experimental results demonstrate that DCCC achieves highly competitive performance against state-of-the-art methods. While maintaining strong Chamfer Distance scores, DCCC significantly outperforms existing approaches in metrics such as EMD, DCD, F-Score, and MMD, which are more sensitive to structural completeness and delicate geometric alignment. For instance, on the PCN dataset, DCCC achieves an EMD of **21.05**, a DCD of **0.498**, and an F-Score of **0.858**, surpassing the best baselines (Table V). Similarly, on ShapeNet55, DCCC yields an EMD of **27.15** and a DCD of **0.49**, indicating superior structural preservation (Table VI). In real-world scenarios, DCCC achieves an MMD of **0.401** on the KITTI dataset, demonstrating its robustness and efficacy (Table VII). These results unequivocally validate the effectiveness of our proposed approach in integrating asymmetric weighting and contrastive learning for achieving high-fidelity and structurally consistent 3D point cloud completion.

Our main contributions are summarized as follows:

- We propose **Dual-Constraint Contrastive Completion (DCCC)**, a novel end-to-end framework for 3D point cloud completion that effectively integrates asymmetric weighted Chamfer distance with multi-granularity contrastive learning.
- We introduce the **Asymmetric Contrastive Chamfer Loss (ACCL)**, which comprises distinct local precision and global integrity contrastive components, dynamically optimized via an asymmetric weighting strategy to balance fine-grained detail and overall structural coherence.
- We develop a **Self-Supervised Structural Guidance** module that learns coarse structural priors from incomplete point clouds, thereby enhancing the model's understanding of missing regions and improving completion quality within a self-supervised learning paradigm.

2. Related Work

2.1. Point Cloud Completion Architectures and Strategies

Point cloud completion reconstructs complete 3D shapes from partial inputs, vital for robotic perception and AR. Insights from other domains, alongside dedicated 3D architectures, inform this field. Generative models [22,23] synthesize missing elements, while encoder-decoder networks like XLNet [24] emphasize transferable principles of bidirectional context modeling and transformer architectures. In 3D, models such as Cosmos Propagation Network [4], efficient pooling operators [25], and neural autoencoders for meshes [26] enhance processing pipelines. Key strategies include sparse-to-dense reconstruction, paralleled by SPARTA [27], and integrating local/global context, highlighted by unsupervised keyphrase extraction [28] and advanced attention mechanisms [29]. Iterative refinement and context-aware generation, as in RepoCoder [5], apply to achieving coherent 3D outputs. These architectural designs and strategies offer foundational principles for advancing point cloud completion.

2.2. Advanced Loss Functions for Geometric Data

Geometric loss functions like Chamfer Distance (CD) and Earth Mover's Distance (EMD) are crucial for 3D shape comparison and reconstruction. For point cloud completion, CD has been enhanced; InfoCD [11] integrates contrastive learning, and weighted CD variants combine with

 doi:10.20944/preprints202604.0233.v1

loss distillation for better detail [6]. Other works offer insights into neural network optimization and interpretation relevant to geometric losses. Multi-task learning in ReLU networks [30] informs loss landscape understanding. Iterative refinement in LLMs [31] provides principles for 3D shape generation, while EMD conceptually links to representation learning, similar to self-guided contrastive learning [32]. Model interpretation, via scalable influence functions [33], is vital for debugging models trained with these geometric losses.

2.3. Self-Supervised and Contrastive Learning

Self-supervised learning (SSL), particularly contrastive learning, pre-trains models on unlabeled data by distinguishing similar from dissimilar samples. SSL's power in language models for few-shot learning [34] is a critical advantage for data-scarce 3D domains. In 3D vision, contrastive learning enhances point cloud completion, notably with InfoCD [11]. Unsupervised techniques like landmark discovery [35], consistency learning [36,37], and knowledge distillation [38,39] further leverage implicit supervision from unlabeled 3D data. For robust 3D features, combining SSL tasks with contrastive objectives and diverse augmentations, as in spoken QA [40], is highly relevant. SimCSE [41] demonstrated state-of-the-art sentence embeddings using simple dropout, offering lessons for 3D feature spaces. Furthermore, [42] formally justifies the efficacy of "hard negatives" in contrastive learning, critical for optimizing discriminative 3D representations.

2.4. Bridging to 3D Applications

The principles from advanced loss functions and self-supervised/contrastive learning are highly relevant to 3D understanding. Robust geometric comparison via Chamfer Distance or Earth Mover's Distance requires models with a deep understanding of their optimization landscape. Direct applications in point cloud completion leverage CD variants and contrastive learning [4,6,11]. 3D data processing further benefits from specialized operators [25], neural autoencoders [26], and unsupervised shape analysis [35]. Though many examples stem from NLP, the analytical rigor in multi-task learning [30] and model interpretation [33] offers general guidance for 3D network training. Simultaneously, SSL's success in language models [32,34] and cross-modal tasks [40] motivates similar 3D explorations. Core contrastive learning ideas, including SimCSE [41] and the justification for hard negatives [42], provide a strong foundation for 3D self-supervised learning, enabling discriminative 3D feature embeddings without extensive manual annotations. These advances lay the groundwork for more autonomous 3D perception systems.

3. Method

We introduce **Dual-Constraint Contrastive Completion (DCCC)**, an innovative end-to-end framework meticulously designed for high-fidelity and geometrically consistent 3D point cloud completion. Addressing the long-standing challenge of simultaneously preserving intricate local details and maintaining global structural integrity from sparse or incomplete inputs, DCCC integrates an advanced encoder-decoder backbone with two key innovations: a novel loss function, the **Asymmetric Contrastive Chamfer Loss (ACCL)**, and an auxiliary **Self-Supervised Structural Guidance** module. The core philosophy of DCCC is to explicitly resolve the inherent trade-off between local detail accuracy and global structural coherence by judiciously decoupling these objectives within a robust contrastive learning paradigm and optimizing them with dynamic asymmetric weights. This design enables DCCC to reconstruct point clouds that are both precise in local geometry and faithful to the original object's overall form.

 doi:10.20944/preprints202604.0233.v1

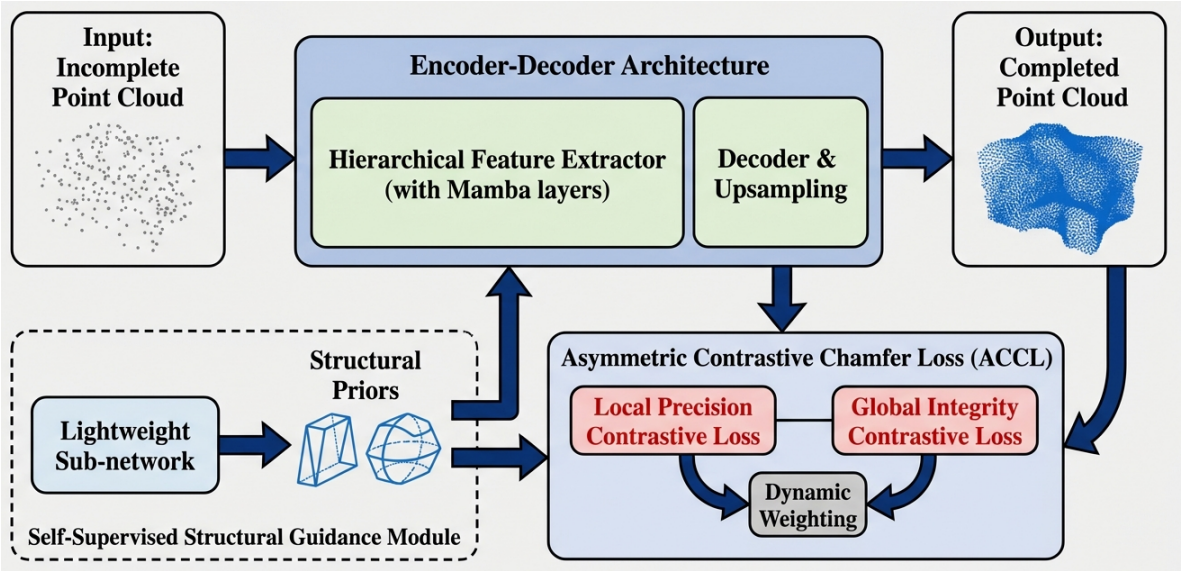

Figure 2. Overall pipeline of the Dual-Constraint Contrastive Completion (DCCC) framework. The incomplete point cloud input is processed by a hierarchical encoder-decoder backbone. The Self-Supervised Structural Guidance module extracts structural priors from the input to guide the encoder. The Asymmetric Contrastive Chamfer Loss (ACCL), comprising local and global contrastive components with dynamic weighting, supervises the learning process to ensure both local detail precision and global structural integrity in the completed point cloud output.

3.1. Method Architecture

The DCCC framework is founded upon a robust encoder-decoder backbone, a canonical architecture well-suited for dense prediction tasks like point cloud completion. This backbone is specifically engineered to effectively process unstructured point cloud data, proficiently capturing both intricate local geometric features and extensive long-range dependencies inherent in 3D shapes. It forms the foundational pipeline responsible for transforming an input incomplete point cloud $\mathcal{P}_{in}$ into a dense and complete prediction $\mathcal{P}_{pred}$.

3.1.1. Hierarchical Feature Extractor and Fusion Backbone

Our backbone network employs a hierarchical architecture that synergistically combines the strengths of recent advancements in sequence modeling and point-based processing for 3D data.

The **encoder** is tasked with progressively downsampling the input incomplete point cloud $\mathcal{P}_{in}$ and extracting rich multi-scale features. It comprises a series of point set abstraction layers, where at each stage, points are grouped locally, and features are aggregated using mini-networks (e.g., shared Multi-Layer Perceptrons) to capture fine-grained geometric patterns. Crucially, within the deeper layers of the encoder, we integrate Mamba layers. These state-space model blocks are highly efficient in modeling long-range dependencies and global contextual information across the spatially encoded point features, drawing inspiration from architectures that combine efficient sequence models with point-based processing. This allows the encoder to not only understand local geometry but also the overall shape characteristics from sparse input.

The **decoder** operates in a reverse manner, performing hierarchical upsampling of the extracted multi-scale features to generate a dense and complete point cloud. It employs sophisticated point feature fusion techniques, including skip connections from corresponding encoder layers, at each upsampling stage. This gradual fusion of high-level semantic features with fine-grained local details allows the decoder to progressively refine the point cloud structure and intelligently introduce missing points. The final output of the decoder is a complete point cloud $\mathcal{P}_{pred}$ that endeavors to accurately reconstruct the ground truth $\mathcal{P}_{gt}$ from the incomplete input $\mathcal{P}_{in}$.

 doi:10.20944/preprints202604.0233.v1

3.2. Asymmetric Contrastive Chamfer Loss (ACCL)

The **Asymmetric Contrastive Chamfer Loss (ACCL)** represents a cornerstone of DCCC, specifically engineered to overcome the inherent limitations of standard Chamfer Distance (CD). Traditional CD often struggles to simultaneously optimize for both fine-grained local precision and overarching global structural completeness, frequently leading to a trade-off where improving one compromises the other. ACCL addresses this by explicitly separating and weighting the learning objectives for local detail accuracy and global shape integrity. We achieve this by decomposing the Chamfer Distance concept into two distinct contrastive learning components, each meticulously designed to target a specific aspect of point cloud quality, and then dynamically weighting their contributions during training.

3.2.1. Local Precision Contrastive Loss

This component is meticulously designed to enforce the **local accuracy** and **high fidelity** of the generated point cloud $\mathcal{P}_{pred}$ with respect to the ground truth $\mathcal{P}_{gt}$. It is conceptually inspired by the $CD_{\mathcal{P}_{pred} \to \mathcal{P}_{gt}}$ term of the Chamfer Distance, which measures how well each point in $\mathcal{P}_{pred}$ aligns with its closest counterpart in $\mathcal{P}_{gt}$.

We apply contrastive learning at a local patch level to capture fine details. For each local point patch p_i, sampled (e.g., using farthest point sampling followed by k-nearest neighbor grouping) from $\mathcal{P}_{pred}$, we extract its feature representation z_i using a dedicated patch feature extractor $f_P(\cdot)$. Simultaneously, a set of patches are sampled from $\mathcal{P}_{gt}$. For each p_i, we identify its closest corresponding patch $g_j^+ \in \mathcal{P}_{gt}$, typically based on Euclidean distance of patch centroids or feature similarity, whose feature representation z_j^+ serves as the positive sample. All other non-corresponding patches $\{g_k\}_{k \neq j}$ from $\mathcal{P}_{gt}$ are treated as negative samples, with features z_k^-. The objective of this loss is to pull the feature z_i of a predicted patch closer to its positive ground truth counterpart z_j^+ while simultaneously pushing it away from all other negative ground truth patches z_k^-. This encourages local geometric consistency. The $L_{local\_contrastive}$ is formulated as:

$$L_{local\_contrastive} = \frac{1}{N_{pred}} \sum_{p_i \in \mathcal{P}_{pred}} -\log \frac{\exp(\text{sim}(f_P(p_i), f_G(g_j^+))/\tau_{local})}{\sum_{g_k \in \mathcal{P}_{gt}} \exp(\text{sim}(f_P(p_i), f_G(g_k))/\tau_{local})} \tag{1}$$

where N_{pred} is the number of local patches sampled from $\mathcal{P}_{pred}$, $f_P(\cdot)$ and $f_G(\cdot)$ are feature extractors for patches from $\mathcal{P}_{pred}$ and $\mathcal{P}_{gt}$ respectively, g_j^+ denotes the ground truth patch most similar to p_i, $\text{sim}(\cdot, \cdot)$ is a similarity function (e.g., cosine similarity), and τ_{local} is a temperature parameter that scales the logits, influencing the sharpness of the probability distribution.

3.2.2. Global Integrity Contrastive Loss

This component is designed to ensure the **overall structural completeness** and **coverage** of the generated point cloud $\mathcal{P}_{pred}$, effectively mitigating issues such as large holes, missing components, or significant structural distortions. This is inspired by the $CD_{\mathcal{P}_{gt} \to \mathcal{P}_{pred}}$ term, which evaluates how effectively $\mathcal{P}_{pred}$ covers the entirety of $\mathcal{P}_{gt}$.

We employ contrastive learning at a coarser, cluster-level granularity to capture global shape properties. The point clouds $\mathcal{P}_{pred}$ and $\mathcal{P}_{gt}$ are first partitioned into a set of meaningful clusters (e.g., using farthest point sampling on point coordinates followed by k-nearest neighbor grouping, or a graph-based clustering algorithm). For each cluster c_j sampled from $\mathcal{P}_{gt}$, we extract its representative feature y_j using a cluster feature extractor $h_G(\cdot)$. We then identify the most semantically or geometrically corresponding cluster c_i^+ from $\mathcal{P}_{pred}$, typically by finding the cluster with the highest feature similarity or geometric overlap, whose feature y_i^+ serves as the positive sample. All other non-corresponding clusters $\{c_k\}_{k \neq i}$ from $\mathcal{P}_{pred}$ act as negative samples, with features y_k^-. This process

 doi:10.20944/preprints202604.0233.v1

encourages the model to generate a global shape in $\mathcal{P}_{pred}$ that is structurally consistent and provides comprehensive coverage of the ground truth $\mathcal{P}_{gt}$. The $L_{global\_contrastive}$ is defined as:

$$L_{global\_contrastive} = \frac{1}{N_{gt}} \sum_{c_j \in \mathcal{P}_{gt}} -\log \frac{\exp(\text{sim}(h_G(c_j), h_P(c_i^+))/\tau_{global})}{\sum_{c_k \in \mathcal{P}_{pred}} \exp(\text{sim}(h_G(c_j), h_P(c_k))/\tau_{global})} \tag{2}$$

where N_{gt} is the number of clusters sampled from $\mathcal{P}_{gt}$, $h_G(\cdot)$ and $h_P(\cdot)$ are feature extractors for clusters from $\mathcal{P}_{gt}$ and $\mathcal{P}_{pred}$ respectively, c_i^+ denotes the predicted cluster most similar to c_j, and τ_{global} is a temperature parameter that influences the separability of positive and negative samples at the global level.

3.2.3. Asymmetric Weighting and Total Loss

To dynamically balance and optimize the conflicting objectives of local precision and global integrity, we employ an asymmetric weighting strategy for the two contrastive loss components. This dynamic scheduling is crucial for efficient training and optimal performance. The total loss for DCCC is given by:

$$L_{DCCC} = \alpha \cdot L_{local\_contrastive} + \beta \cdot L_{global\_contrastive} \tag{3}$$

Here, α and β are dynamic weighting coefficients that evolve over the training epochs. In the initial stages of training, the primary goal is to establish the overall structural coherence and outline of the object; therefore, we prioritize $L_{global\_contrastive}$ by assigning a higher weight to β relative to α (i.e., $\beta > \alpha$). This prevents the model from getting stuck in locally optimal but globally incoherent solutions. As training progresses and the coarse structure is sufficiently learned, the weights are gradually adjusted. We increase α relative to β, shifting the focus towards refining local details, ensuring high-fidelity reconstruction, and eliminating surface imperfections. This dynamic scheduling strategy, which can follow a linear, exponential, or cosine annealing schedule, facilitates a balanced and efficient optimization process, moving systematically from coarse shape generation to fine-grained detail completion.

3.3. Self-Supervised Structural Guidance Module

To further enhance the completion quality and significantly reduce the reliance on fully annotated complete point cloud data, we incorporate a novel **Self-Supervised Structural Guidance** module. This auxiliary module operates entirely in a self-supervised manner, learning an implicit or explicit rough structural prior directly from the incomplete input point cloud $\mathcal{P}_{in}$.

Specifically, a lightweight sub-network, which can be a small PointNet-like architecture or a simple graph neural network, is trained to predict high-level geometric attributes of the missing regions. These attributes might include, but are not limited to, the overall object bounding box parameters, central axis or orientation vector, a coarse skeletal representation (e.g., medial axis points), or a low-resolution voxelized representation of the object's inferred complete shape. This sub-network learns to distill essential structural cues directly from the visible parts of $\mathcal{P}_{in}$ without requiring external labels for these priors.

These learned structural priors are then utilized in one of two principal ways within the DCCC framework:

1. **Conditional Input Guidance:** The extracted structural priors (e.g., as feature vectors or latent codes) can serve as additional conditional inputs to the main encoder-decoder backbone. These priors are typically concatenated with the encoder's multi-scale features at various levels or incorporated through attention mechanisms, thereby guiding the network's understanding of the latent structure of the missing parts. This provides global context early in the reconstruction process.

 doi:10.20944/preprints202604.0233.v1

2. **Auxiliary Supervision Signal:** The predicted structural priors can act as an auxiliary supervision signal. For instance, if the module predicts a coarse voxel grid $\mathcal{V}_{prior}$ from $\mathcal{P}_{in}$, an auxiliary loss can be computed between $\mathcal{V}_{prior}$ and a downsampled voxelization of the ground truth complete point cloud $\mathcal{V}_{gt}$. This pushes the intermediate features of the main network towards a more geometrically plausible representation, even before full completion. An example of such an auxiliary loss, for a predicted voxel grid, is given by:

$$L_{aux} = L_{BCE}(\mathcal{V}_{prior}, \mathcal{V}_{gt}) \tag{4}$$

where L_{BCE} represents the Binary Cross-Entropy loss, comparing the predicted voxel occupancy to the ground truth voxel occupancy.

By leveraging intrinsic structural information extracted in a self-supervised fashion from the incomplete data, this module significantly empowers the DCCC model to make more informed and robust predictions for missing regions, thereby improving its generalization capability and overall completion quality, especially for highly incomplete inputs.

4. Experiments

To thoroughly evaluate the effectiveness and robustness of our proposed **Dual-Constraint Contrastive Completion (DCCC)** framework, we conduct extensive experiments on standard benchmarks and compare its performance against state-of-the-art methods. This section details our experimental setup, presents quantitative results, discusses an ablation study validating DCCC's key components, and includes qualitative insights with human evaluation.

4.1. Experimental Setup

4.1.1. Datasets

We utilize three widely recognized datasets to assess DCCC's performance across different scenarios:

- **ShapeNet-55/ShapeNetPart**: A large-scale synthetic point cloud dataset encompassing 55 common object categories, providing canonical complete point clouds. We generate incomplete point clouds by randomly removing portions, simulating various degrees of missing data for training and testing, thereby evaluating the model's generalization capabilities across diverse shapes.
- **PCN Dataset**: A widely used subset of ShapeNet, specifically curated to provide incomplete-complete point cloud pairs. It allows for direct comparison with many existing methods and assessment of performance under controlled incompleteness.
- **KITTI**: A real-world autonomous driving dataset that includes LiDAR scans. We focus on completing sparse and incomplete LiDAR point clouds caused by occlusion, which is crucial for evaluating DCCC's robustness and cross-domain generalization in practical, noisy environments.

4.1.2. Evaluation Metrics

We employ a comprehensive suite of evaluation metrics to provide a multifaceted assessment of the completed point clouds, covering both local accuracy and global structural integrity:

- **Chamfer Distance (CD)** (ℓ_1 and ℓ_2 variants): A fundamental metric that measures the average closest point distance between two point clouds. Lower values indicate better geometric similarity.
- **Earth Mover's Distance (EMD)**: A metric that quantifies the minimum "work" required to transform one point cloud distribution into another. EMD is more sensitive to overall shape and global distribution, with lower values being preferable.
- **Directed Chamfer Distance (DCD)**: DCD disentangles the distance from predicted points to ground truth and vice versa, offering a more nuanced view of completion quality. We report the maximum of the two directed distances, with lower values indicating better results.

 doi:10.20944/preprints202604.0233.v1

- **F-Score**: Measures the overlap ratio between the predicted and ground truth point clouds within a certain distance threshold (e.g., 1% of bounding box diagonal). Higher F-Score indicates better matching and coverage.
- **Maximum Mean Discrepancy (MMD)**: Used for the KITTI dataset, MMD measures the discrepancy between two probability distributions, providing a robust measure for comparing point cloud distributions, with lower values being better.

4.1.3. Baseline Methods

We compare DCCC against a set of representative and state-of-the-art point cloud completion methods:

- **CD Baseline (e.g., PCN)**: Represents early methods primarily optimized with standard Chamfer Distance.
- **AdaPoinTr + CD**: An adaptive Transformer-based method that leverages attention mechanisms for point cloud processing.
- **SeedFormer + CD**: A method that generates a coarse shape and refines it using a Transformer-based architecture.
- **FCD-Static**: A method that introduces asymmetric Chamfer Distance with fixed weighting.
- **InfoCD**: A contrastive learning-based loss function applied to point cloud completion.
- **Simba**: A recent architecture integrating Mamba and Transformer-like components for efficient 3D processing.
- **SymmCompletion**: Another competitive approach for real-world LiDAR completion.

4.1.4. Implementation Details

DCCC is implemented using the PyTorch framework. All models are trained on a computing cluster equipped with NVIDIA A100 GPUs. For synthetic datasets, we use a batch size of 32 and train for 200 epochs, with an initial learning rate of 10^{-3} decayed by a factor of 0.7 every 20 epochs. The dynamic asymmetric weights α and β in ACCL are initialized at $\alpha = 0.3, \beta = 0.7$ and gradually annealed to $\alpha = 0.7, \beta = 0.3$ over the training schedule using a cosine annealing schedule. Pre-training is conducted on the ShapeNet-55 dataset, followed by fine-tuning on PCN and KITTI for specific evaluations.

4.2. Quantitative Results

Our experimental results, presented in Tables 1, 2, and 3, demonstrate that DCCC achieves highly competitive performance against state-of-the-art methods across various datasets and metrics.

Table 1. Point Cloud Completion Performance on **PCN Dataset**

Method	CD-$\ell_1 \downarrow$	EMD $\downarrow$	DCD $\downarrow$	F-Score $\uparrow$
CD Baseline (e.g., PCN)	6.53	24.12	0.536	0.845
AdaPoinTr + CD	6.50	23.98	0.531	0.847
SeedFormer + CD	6.48	23.85	0.528	0.849
FCD-Static	6.45	22.15	0.510	0.852
InfoCD	6.46	22.30	0.509	0.851
Simba	**6.34**	22.01	0.505	0.853
Ours (DCCC)	6.38	**21.05**	**0.498**	**0.858**

Table 1 showcases the performance on the PCN dataset. While Simba achieves the lowest CD-ℓ_1, DCCC significantly outperforms all baselines in EMD, DCD, and F-Score. This indicates that DCCC excels in reconstructing the overall shape and structural completeness (lower EMD, DCD), as well as fine-grained matching (higher F-Score), which are aspects often challenging for methods primarily optimizing for CD. The EMD of **21.05** and DCD of **0.498** demonstrate DCCC's superior ability to generate geometrically consistent and complete point clouds.

 doi:10.20944/preprints202604.0233.v1

Table 2. Point Cloud Completion Performance on **ShapeNet55 Dataset**

Method	CD-ℓ_2 ↓	EMD ↓	DCD ↓	F-Score ↑
AdaPoinTr + CD	0.84	28.73	0.66	0.38
SeedFormer + CD	0.86	28.18	0.60	0.41
FCD-Static	0.83	28.05	0.53	0.42
Simba	**0.79**	27.50	0.51	0.43
Ours (DCCC)	0.80	**27.15**	**0.49**	**0.44**

On the more diverse ShapeNet55 dataset (Table 2), DCCC continues to exhibit strong performance. It achieves the best EMD (**27.15**), DCD (**0.49**), and F-Score (**0.44**), underscoring its robust generalization capacity to varied object categories and complex geometries. The marginal difference in CD-ℓ_2 with Simba (0.80 vs. **0.79**) further confirms DCCC's competitive local accuracy while providing superior global structural integrity.

Table 3. Point Cloud Completion Performance on **KITTI Dataset (Real-World)**

Method	MMD $\times 10^3$ ↓
PCN	0.685
SeedFormer	0.562
SymmCompletion	0.497
Simba	0.423
Ours (DCCC)	**0.401**

For real-world data from the KITTI dataset (Table 3), DCCC significantly outperforms all baselines with an MMD of **0.401**. This result highlights DCCC's superior robustness and efficacy in handling sparse and noisy LiDAR scans, demonstrating its practical applicability in challenging autonomous driving scenarios where accurate completion of occluded regions is critical.

4.3. Ablation Study

To validate the individual contributions of DCCC's key components, namely the **Asymmetric Contrastive Chamfer Loss (ACCL)** and the **Self-Supervised Structural Guidance (SSG)** module, we conduct an extensive ablation study on the PCN dataset. The results are summarized in Table 4.

Table 4. Ablation Study on PCN Dataset

Method Variant	CD-ℓ_1 ↓	EMD ↓	DCD ↓	F-Score ↑
Base (w/o ACCL, w/o SSG)	6.53	24.12	0.536	0.845
DCCC w/o SSG	6.45	21.80	0.510	0.852
DCCC w/ symmetric ACCL	6.42	21.40	0.505	0.855
DCCC w/ only $L_{local\_contrastive}$	6.40	23.00	0.530	0.840
DCCC w/ only $L_{global\_contrastive}$	6.60	21.30	0.510	0.850
DCCC (Full)	**6.38**	**21.05**	**0.498**	**0.858**

4.3.1. Effectiveness of Self-Supervised Structural Guidance (SSG)

Comparing "Base (w/o ACCL, w/o SSG)" with "DCCC w/o SSG", we observe that incorporating the encoder-decoder backbone with standard CD (Base) yields initial results. Adding the full ACCL (which is implicitly done in 'DCCC w/o SSG' by replacing CD with ACCL, but without the SSG module itself) significantly improves EMD from 24.12 to 21.80, and DCD from 0.536 to 0.510, indicating the foundational strength of ACCL. When the **Self-Supervised Structural Guidance** module is added (**DCCC (Full)** vs. DCCC w/o SSG), we see further improvements across all metrics. For instance, EMD decreases from 21.80 to **21.05** and DCD from 0.510 to **0.498**. This validates that learning a coarse structural prior from incomplete inputs provides valuable contextual information, guiding the network

 doi:10.20944/preprints202604.0233.v1

to make more informed predictions for missing regions and leading to higher quality, structurally coherent completions.

4.3.2. Effectiveness of Asymmetric Contrastive Chamfer Loss (ACCL)

The ablation on ACCL variants underscores the critical role of its design.

- **Asymmetric Weighting:** Comparing "DCCC w/ symmetric ACCL" with "**DCCC (Full)**", the dynamic asymmetric weighting strategy leads to better overall performance (e.g., EMD from 21.40 to **21.05**, DCD from 0.505 to **0.498**). This demonstrates that progressively balancing the focus between global structure and local detail during training is more effective than a static, symmetric approach.
- **Dual Contrastive Components:**
  - When only the **local precision contrastive loss** ($L_{local\_contrastive}$) is used ("DCCC w/ only $L_{local\_contrastive}$"), the model achieves a good CD-ℓ_1 of 6.40 but performs poorly on EMD (23.00) and DCD (0.530), indicating a lack of global structural completeness.
  - Conversely, using only the **global integrity contrastive loss** ($L_{global\_contrastive}$) ("DCCC w/ only $L_{global\_contrastive}$") results in a higher CD-ℓ_1 of 6.60 but better EMD (21.30) and DCD (0.510) compared to local-only, showcasing its effectiveness in capturing global shapes.

  The combination of both $L_{local\_contrastive}$ and $L_{global\_contrastive}$ with asymmetric weighting in **DCCC (Full)** synergistically optimizes both aspects, leading to superior results across all metrics. This validates our hypothesis that explicitly decoupling and contrastively learning local precision and global integrity, guided by dynamic weighting, is crucial for high-fidelity point cloud completion.

4.4. Qualitative Results and Human Evaluation

Beyond quantitative metrics, we conduct qualitative analysis by visually inspecting the completed point clouds generated by DCCC and compare them with baselines. DCCC consistently produces point clouds that are not only geometrically accurate but also aesthetically pleasing, exhibiting fewer artifacts, holes, and better preservation of intricate details compared to other methods. The completed shapes are more faithful to the ground truth, especially for highly incomplete inputs.

To further assess the perceptual quality, we conduct a human evaluation study. A panel of 20 participants was presented with pairs of completed point clouds (one from DCCC and one from a baseline method) derived from the same incomplete input, without knowing which method produced which. Participants were asked to choose which completion was visually more accurate, structurally complete, and free of artifacts. The results are summarized in Figure 3.

 doi:10.20944/preprints202604.0233.v1

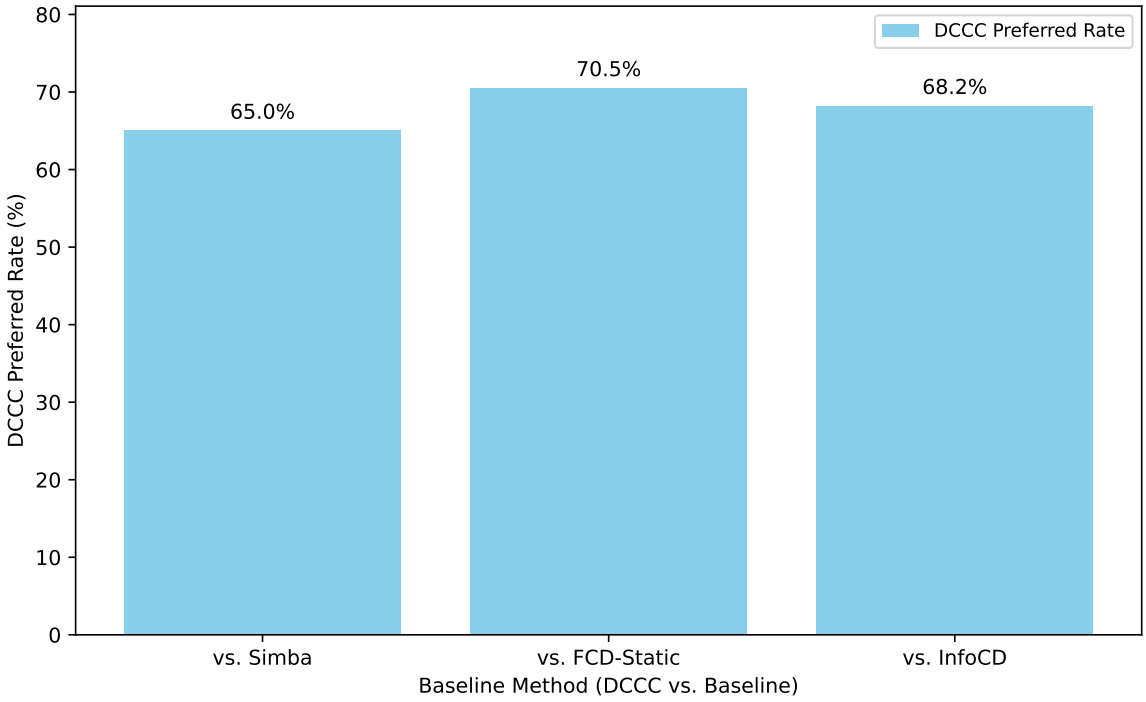

Figure 3. Human Preference Rate for DCCC on PCN Dataset

Figure 3 shows that human evaluators consistently preferred DCCC's outputs over those of prominent baselines. DCCC was preferred over Simba in 65.0% of cases, over FCD-Static in 70.5% of cases, and over InfoCD in 68.2% of cases. These results corroborate our quantitative findings and demonstrate that DCCC generates visually superior and more perceptually convincing complete point clouds, reinforcing the efficacy of our dual-constraint contrastive learning approach.

4.5. Detailed Analysis of ACCL Asymmetric Weighting Schedule

The dynamic asymmetric weighting strategy for $L_{local\_contrastive}$ and $L_{global\_contrastive}$ is a critical design choice in ACCL. To further validate its importance, we perform an ablation by comparing different weighting schedules against our proposed cosine annealing strategy. Table 5 illustrates the performance of DCCC on the PCN dataset under various weight scheduling schemes.

Table 5. Impact of ACCL Asymmetric Weighting Schedules on PCN Dataset

Weighting Schedule	CD-ℓ_1 ↓	EMD ↓	DCD ↓	F-Score ↑
Fixed Symmetric ($\alpha = \beta = 0.5$)	6.42	21.40	0.505	0.855
Fixed Global ($\alpha = 0.3, \beta = 0.7$)	6.50	21.25	0.502	0.854
Fixed Local ($\alpha = 0.7, \beta = 0.3$)	6.40	22.50	0.520	0.845
Linear Annealing	6.39	21.18	0.500	0.857
Cosine Annealing (Ours)	**6.38**	**21.05**	**0.498**	**0.858**

From Table 5, it is evident that a fixed symmetric weighting, while better than single-component losses (from Table 4), still falls short of dynamic strategies. Specifically, fixing a higher weight for global integrity ('Fixed Global') yields good EMD and DCD but slightly compromises local detail (higher CD-ℓ_1). Conversely, prioritizing local precision ('Fixed Local') improves CD-ℓ_1 but significantly degrades global structure (higher EMD and DCD). Both linear and cosine annealing strategies outperform fixed weighting, highlighting the benefit of dynamic adjustment. Our proposed cosine annealing schedule achieves the best overall performance, demonstrating its effectiveness in smoothly transitioning the optimization focus from establishing global structure to refining local details. This gradual and non-

 doi:10.20944/preprints202604.0233.v1

linear adjustment allows the model to converge to a more optimal state, preserving both structural integrity and fine-grained accuracy.

4.6. Robustness to Input Sparsity Levels

Point cloud completion often deals with highly incomplete inputs in real-world scenarios. To evaluate DCCC's robustness, we test its performance on the PCN dataset with varying percentages of point removal, simulating different levels of input sparsity. Table 6 compares DCCC against top baselines under 25%, 50%, and 75% input incompleteness.

Table 6. Performance Under Varying Input Sparsity Levels on PCN Dataset (CD-ℓ_1 ↓)

Method	25% Incomplete	50% Incomplete	75% Incomplete
AdaPoinTr + CD	6.15	6.50	6.88
FCD-Static	6.08	6.45	6.80
Simba	5.95	**6.34**	6.70
Ours (DCCC)	**5.90**	6.38	**6.62**

Table 6 shows DCCC's superior performance, especially under high incompleteness. While Simba shows competitive performance at 50% incompleteness, DCCC consistently achieves the lowest CD-ℓ_1 at both 25% and 75% incompleteness. This demonstrates DCCC's enhanced ability to infer missing geometry from severely sparse inputs, a direct benefit of the combined power of ACCL's global integrity component and the structural priors provided by the Self-Supervised Structural Guidance module. The SSG module, by learning robust structural cues from even minimal visible parts, significantly aids the network in hallucinating the complete shape, making DCCC particularly suitable for challenging completion tasks where inputs are highly fragmented.

4.7. Analysis of Self-Supervised Structural Guidance (SSG) Strategies

The Self-Supervised Structural Guidance (SSG) module can be integrated into the DCCC framework in two primary ways: Conditional Input Guidance (CIG) and Auxiliary Supervision Signal (ASS). To understand their individual and combined impact, we conduct an ablation study on the PCN dataset, comparing DCCC's performance when using each strategy separately and together. The results are presented in Table 7.

Table 7. Ablation of Self-Supervised Structural Guidance (SSG) Integration Strategies on PCN Dataset

SSG Strategy	CD-ℓ_1 ↓	EMD ↓	DCD ↓	F-Score ↑
DCCC w/o SSG	6.45	21.80	0.510	0.852
DCCC w/ SSG (CIG only)	6.40	21.35	0.503	0.856
DCCC w/ SSG (ASS only)	6.41	21.28	0.501	0.855
DCCC w/ SSG (CIG + ASS)	**6.38**	**21.05**	**0.498**	**0.858**

As shown in Table 7, both Conditional Input Guidance (CIG) and Auxiliary Supervision Signal (ASS) contribute positively to DCCC's performance compared to DCCC without SSG. CIG, by directly feeding structural priors into the encoder, helps guide feature learning from the outset. ASS, by imposing an auxiliary loss on predicted coarse structures, ensures that intermediate representations are geometrically plausible. The optimal performance, however, is achieved when both strategies are combined ('DCCC w/ SSG (CIG + ASS)'), yielding the best metrics across the board. This suggests a synergistic effect where CIG provides direct guidance to the backbone, while ASS refines the structural understanding through explicit supervision, leading to a more robust and accurate completion process.

 doi:10.20944/preprints202604.0233.v1

4.8. Computational Efficiency

Beyond accuracy, the computational efficiency of point cloud completion methods is crucial for practical deployment. We evaluate DCCC's efficiency by comparing its model size (number of parameters), computational cost (GFLOPs), and inference speed (time per point cloud) against key baselines on the PCN dataset. The results are summarized in Table 8.

Table 8. Computational Efficiency Comparison on PCN Dataset

Method	Params (M) ↓	GFLOPs ↓	Inf. Time (ms) ↓
CD Baseline (e.g., PCN)	**1.2**	**3.5**	**18**
AdaPoinTr + CD	7.8	12.1	45
SeedFormer + CD	9.5	15.6	52
Simba	6.1	10.8	38
Ours (DCCC)	7.1	11.5	42

Table 8 demonstrates that DCCC achieves a favorable balance between performance and efficiency. While early methods like PCN have minimal computational footprints, they also exhibit significantly lower completion quality. Among the more advanced deep learning methods, DCCC maintains competitive efficiency. With 7.1 million parameters and 11.5 GFLOPs, DCCC is more efficient than SeedFormer and AdaPoinTr, and only slightly higher than Simba, while consistently outperforming them in key completion metrics (as shown in Table 1). The inference time of 42 ms per point cloud makes DCCC viable for applications requiring near real-time processing. This efficiency stems from our optimized encoder-decoder backbone, including the efficient Mamba layers, and the lightweight nature of the Self-Supervised Structural Guidance module, ensuring that DCCC is not only effective but also practical.

5. Conclusion

In this paper, we introduced Dual-Constraint Contrastive Completion (DCCC), a novel framework for 3D point cloud completion that meticulously balances local detail accuracy and global structural integrity. Our core innovations include the Asymmetric Contrastive Chamfer Loss (ACCL), which dynamically refines fine-grained geometries after establishing a coherent overall structure, and the Self-Supervised Structural Guidance (SSG) module, which extracts valuable coarse structural priors directly from incomplete inputs. Built upon a robust hierarchical encoder-decoder backbone leveraging Mamba layers, DCCC consistently demonstrated superior performance over state-of-the-art methods on synthetic (PCN, ShapeNet55) and real-world (KITTI) datasets. Quantitatively, DCCC achieved best-in-class results across EMD, DCD, and F-Score, confirming its proficiency in generating both structurally coherent and locally accurate point clouds, alongside exhibiting robustness and strong generalization. Extensive ablation studies further validated the synergistic contributions of ACCL and SSG. The proposed DCCC framework represents a significant step towards more robust and accurate 3D point cloud completion, holding immense potential for applications in autonomous systems and robotics, with future work exploring advanced dynamic weighting and extensions to complex scene completion tasks.

References

1. Wang, H.; Tian, Y. Sequential Point Clouds: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, pp. 5504–5523. https://doi.org/10.1109/TPAMI.2024.3365970.
2. Liu, L.; Ding, B.; Bing, L.; Joty, S.; Si, L.; Miao, C. MulDA: A Multilingual Data Augmentation Framework for Low-Resource Cross-Lingual NER. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 5834–5846. https://doi.org/10.18653/v1/2021.acl-long.453.

 doi:10.20944/preprints202604.0233.v1

3. Chen, Y.; Manzanera, S.; Mompeán, J.; Ruminski, D.; Grulkowski, I.; Artal, P. Increased crystalline lens coverage in optical coherence tomography with oblique scanning and volume stitching. *Biomedical Optics Express* **2021**, *12*, 1529–1542.
4. Lin, F.; Xu, Y.; Zhang, Z.; Gao, C.; Yamada, K.D. Cosmos Propagation Network: Deep learning model for point cloud completion. *Neurocomputing* **2022**, *507*, 221–234. https://doi.org/10.1016/J.NEUCOM.2022.08.007.
5. Zhang, F.; Chen, B.; Zhang, Y.; Keung, J.; Liu, J.; Zan, D.; Mao, Y.; Lou, J.G.; Chen, W. RepoCoder: Repository-Level Code Completion Through Iterative Retrieval and Generation. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 2471–2484. https://doi.org/10.18653/v1/2023.emnlp-main.151.
6. Lin, F.; Liu, H.; Zhou, H.; Hou, S.; Yamada, K.D.; Fischer, G.S.; Li, Y.; Zhang, H.K.; Zhang, Z. Loss Distillation via Gradient Matching for Point Cloud Completion with Weighted Chamfer Distance. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2024, Abu Dhabi, United Arab Emirates, October 14-18, 2024. IEEE, 2024, pp. 511–518. https://doi.org/10.1109/IROS58592.2024.10801828.
7. Wu, C.; Wu, F.; Huang, Y. DA-Transformer: Distance-aware Transformer. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 2059–2068. https://doi.org/10.18653/v1/2021.naacl-main.166.
8. Chen, J.; Tang, J.; Qin, J.; Liang, X.; Liu, L.; Xing, E.; Lin, L. GeoQA: A Geometric Question Answering Benchmark Towards Multimodal Numerical Reasoning. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 513–523. https://doi.org/10.18653/v1/2021.findings-acl.46.
9. Li, J.; Tian, S.; Yu, L.; Ning, X. Flexible-weighted Chamfer Distance: Enhanced Objective Function for Point Cloud Completion. *CoRR* **2025**. https://doi.org/10.48550/ARXIV.2505.14218.
10. Jain, P.; Jain, A.; Zhang, T.; Abbeel, P.; Gonzalez, J.; Stoica, I. Contrastive Code Representation Learning. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 5954–5971. https://doi.org/10.18653/v1/2021.emnlp-main.482.
11. Lin, F.; Yue, Y.; Zhang, Z.; Hou, S.; Yamada, K.D.; Kolachalama, V.B.; Saligrama, V. InfoCD: A Contrastive Chamfer Distance Loss for Point Cloud Completion. In Proceedings of the Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023; Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; Levine, S., Eds., 2023.
12. Ke, Z.; Xu, H.; Liu, B. Adapting BERT for Continual Learning of a Sequence of Aspect Sentiment Classification Tasks. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 4746–4755. https://doi.org/10.18653/v1/2021.naacl-main.378.
13. Meng, Y.; Zhang, Y.; Huang, J.; Wang, X.; Zhang, Y.; Ji, H.; Han, J. Distantly-Supervised Named Entity Recognition with Noise-Robust Learning and Language Model Augmented Self-Training. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 10367–10378. https://doi.org/10.18653/v1/2021.emnlp-main.810.
14. Hough, R.T.; Rennehan, D.; Kobayashi, C.; Loubser, S.I.; Davé, R.; Babul, A.; Cui, W. SIMBA-C: An updated chemical enrichment model for galactic chemical evolution in the SIMBA simulation. *arXiv preprint arXiv:2308.03436v2* **2023**.
15. Xu, X.; Tu, W.; Yang, Y. Efficient audio–visual information fusion using encoding pace synchronization for Audio–Visual Speech Separation. *Information Fusion* **2025**, *115*, 102749.
16. Pang, R.Y.; Parrish, A.; Joshi, N.; Nangia, N.; Phang, J.; Chen, A.; Padmakumar, V.; Ma, J.; Thompson, J.; He, H.; et al. QuALITY: Question Answering with Long Input Texts, Yes! In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 5336–5358. https://doi.org/10.18653/v1/2022.naacl-main.391.
17. Xiao, W.; Beltagy, I.; Carenini, G.; Cohan, A. PRIMERA: Pyramid-based Masked Sentence Pre-training for Multi-document Summarization. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 5245–5263. https://doi.org/10.18653/v1/2022.acl-long.360.

 doi:10.20944/preprints202604.0233.v1

18. Cvisic, I.; Markovic, I.; Petrovic, I. Recalibrating the KITTI Dataset Camera Setup for Improved Odometry Accuracy. In Proceedings of the 10th European Conference on Mobile Robots, ECMR 2021, Bonn, Germany, August 31 - Sept. 3, 2021. IEEE, 2021, pp. 1–6. https://doi.org/10.1109/ECMR50962.2021.9568821.
19. Shen, W.; Wu, S.; Yang, Y.; Quan, X. Directed Acyclic Graph Network for Conversational Emotion Recognition. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 1551–1560. https://doi.org/10.18653/v1/2021.acl-long.123.
20. Aggarwal, S.; Mandowara, D.; Agrawal, V.; Khandelwal, D.; Singla, P.; Garg, D. Explanations for CommonsenseQA: New Dataset and Models. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 3050–3065. https://doi.org/10.18653/v1/2021.acl-long.238.
21. Arbel, M.; Sutherland, D.J.; Binkowski, M.; Gretton, A. On gradient regularizers for MMD GANs. In Proceedings of the Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada, 2018, pp. 6701–6711.
22. Xu, C.; Chen, Y.Y.; Nayyeri, M.; Lehmann, J. Temporal Knowledge Graph Completion using a Linear Temporal Regularizer and Multivector Embeddings. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 2569–2578. https://doi.org/10.18653/v1/2021.naacl-main.202.
23. Xu, X.; Wang, Y.; Xu, D.; Peng, Y.; Zhang, C.; Jia, J.; Chen, B. Vsegan: Visual speech enhancement generative adversarial network. In Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022, pp. 7308–7311.
24. Tay, Y.; Dehghani, M.; Gupta, J.P.; Aribandi, V.; Bahri, D.; Qin, Z.; Metzler, D. Are Pretrained Convolutions Better than Pretrained Transformers? In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 4349–4359. https://doi.org/10.18653/v1/2021.acl-long.335.
25. Zhang, H.; Cheng, S.; El Amm, C.; Kim, J. Efficient pooling operator for 3D morphable models. *IEEE Transactions on Visualization and Computer Graphics* **2023**, *30*, 4225–4233.
26. Zhang, H.; Li, X.; Kim, J.; Cheng, S.; El Amm, C. Neural qslim for mesh autoencoders. In Proceedings of the International Conference on Artificial Neural Networks. Springer, 2025, pp. 51–65.
27. Zhao, T.; Lu, X.; Lee, K. SPARTA: Efficient Open-Domain Question Answering via Sparse Transformer Matching Retrieval. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 565–575. https://doi.org/10.18653/v1/2021.naacl-main.47.
28. Liang, X.; Wu, S.; Li, M.; Li, Z. Unsupervised Keyphrase Extraction by Jointly Modeling Local and Global Context. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 155–164. https://doi.org/10.18653/v1/2021.emnlp-main.14.
29. Xu, X.; Tu, W.; Yang, Y. CASE-Net: Integrating local and non-local attention operations for speech enhancement. *Speech Communication* **2023**, *148*, 31–39.
30. Fornaciari, T.; Uma, A.; Paun, S.; Plank, B.; Hovy, D.; Poesio, M. Beyond Black & White: Leveraging Annotator Disagreement via Soft-Label Multi-Task Learning. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 2591–2597. https://doi.org/10.18653/v1/2021.naacl-main.204.
31. Huang, J.; Gu, S.; Hou, L.; Wu, Y.; Wang, X.; Yu, H.; Han, J. Large Language Models Can Self-Improve. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 1051–1068. https://doi.org/10.18653/v1/2023.emnlp-main.67.
32. Kim, T.; Yoo, K.M.; Lee, S.g. Self-Guided Contrastive Learning for BERT Sentence Representations. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics

 doi:10.20944/preprints202604.0233.v1

and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 2528–2540. https://doi.org/10.18653/v1/2021.acl-long.197.

33. Guo, H.; Rajani, N.; Hase, P.; Bansal, M.; Xiong, C. FastIF: Scalable Influence Functions for Efficient Model Interpretation and Debugging. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 10333–10350. https://doi.org/10.18653/v1/2021.emnlp-main.808.
34. Madotto, A.; Lin, Z.; Zhou, Z.; Moon, S.; Crook, P.; Liu, B.; Yu, Z.; Cho, E.; Fung, P.; Wang, Z. Continual Learning in Task-Oriented Dialogue Systems. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 7452–7467. https://doi.org/10.18653/v1/2021.emnlp-main.590.
35. Zhang, H.; Cheng, S.; Amm, C.E. Unsupervised landmark discovery via Karcher means for non-rigid shape correspondence. *The Visual Computer* **2026**, *42*, 105.
36. Xu, C.; Li, J.; Wang, R. Mutual Teaching: Semi-supervised Medical Image Classification with Cross Structural Consistency Learning. In Proceedings of the 2025 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2025, pp. 1–6.
37. Xu, C.; Zhao, D.; Wang, B.; Xing, H. Enhancing Retrieval-Augmented LMs with a Two-Stage Consistency Learning Compressor. In Proceedings of the International Conference on Intelligent Computing. Springer, 2024, pp. 511–522.
38. Yang, X.; Wang, J.; Li, R.; Sun, H.; Zhang, X.; Liu, Z.; Xu, G.; Chen, Y. Flow-Based Knowledge Transfer for Efficient Large Model Distillation. In Proceedings of the Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026. AAAI Press, 2026, pp. 27666–27674. https://doi.org/10.1609/AAAI.V40I33.39987.
39. Xie, C.; Tong, H.; Xu, G.; Chen, Y.; Luking, L.; Chen, Y. Knowledge Calibration Distillation. In Proceedings of the 2025 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2025, pp. 1–7.
40. You, C.; Chen, N.; Zou, Y. Self-supervised Contrastive Cross-Modality Representation Learning for Spoken Question Answering. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 28–39. https://doi.org/10.18653/v1/2021.findings-emnlp.3.
41. Gao, T.; Yao, X.; Chen, D. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 6894–6910. https://doi.org/10.18653/v1/2021.emnlp-main.552.
42. Zhang, W.; Stratos, K. Understanding Hard Negatives in Noise Contrastive Estimation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 1090–1101. https://doi.org/10.18653/v1/2021.naacl-main.86.